

Real Estate Market Insights: Zameen.com Listings in Pakistan

Prepared by: Ayesha Jahangir

1. Problem Statement

This analysis explores key drivers of residential property prices in Pakistan using scraped listings data from Zameen.com. The aim is to understand how factors such as location, property type, area, and amenities influence market pricing. Insights generated help investors compare cities, evaluate property value, and make informed real-estate decisions.

2. Dataset Overview

The dataset contains scraped property listings, including city, area, price, property type, bedrooms, bathrooms, and amenity fields. Multiple text inconsistencies and mixed units were present and required cleaning before analysis.

3. Data Cleaning & Preprocessing

Key steps performed:

Removed duplicate entries.

Converted messy "PKR" price strings (Crore/Lakh/Thousand) into numeric PKR values.

Standardized area units (Marla, Kanal, Sq.Yd, Sq.Ft → Sq.Ft).

Cleaned city names and removed whitespace inconsistencies.

Converted bedrooms/bathrooms to numeric values.

Computed Price_per_SqFt for fair comparisons.

4. Missing Values

Core variables such as city, price, area were mostly complete after transformation. Many amenity columns were sparsely filled and used only descriptively.

5. Feature Engineering

Price_per_SqFt: Primary comparison metric.

City_clean: Standardized city labels.

6. Exploratory Analysis Findings

City-Level Insights:

Islamabad: Highest median prices; premium market.

Karachi: Strong price-per-square-foot and demand.

Lahore & Rawalpindi: Upper-mid-range markets.

Faisalabad & Multan: More affordable, value-for-money markets.

General Patterns:

Property prices are right-skewed due to luxury listings.

Most houses have 3–6 bedrooms, matching typical family buyers.

Price_per_SqFt provides more meaningful comparisons than total price.

7. Recommendations

Investors seeking premium markets should target Islamabad and Karachi.

Lahore and Rawalpindi offer balanced price-to-size opportunities.

Value buyers may consider Faisalabad and Multan for lower PPSF.

Platforms should improve listing completeness—amenities are often missing.

8. Limitations & Next Steps

Scraped data snapshot—may not represent full market.

Some units (e.g., Marla size) vary across cities.

Future work: Rental vs sale segmentation, predictive pricing models, neighbourhood-level geographic analysis.

Conclusion

This EDA provides clear insights into Pakistan's real estate pricing dynamics. By cleaning the data and standardizing units, we reveal consistent patterns across cities and property types, enabling investors to make informed decisions using data-driven evidence.